

Instagram, Human Rights, and Ideological Censorship: An In-Depth Analysis

USA, FLORIDA, April 12, 2025

Introduction

Instagram – one of the world’s largest social media platforms – has faced growing scrutiny over allegations of censorship and discrimination against users whose views do not align with the platform’s prevailing content policies or ideology. Critics argue that the company’s content moderation practices sometimes violate freedom of expression and even amount to human rights infringements, particularly in cases involving **conservative, religious, or patriotic viewpoints**. This research examines documented instances where Instagram has **blocked or removed content, banned or shadowbanned users, disabled monetization**, or otherwise suppressed speech, allegedly in a biased manner. It will review notable precedents during the **COVID-19 pandemic**, the **2020 U.S. election and Donald Trump’s campaign**, and the **Russia-Ukraine conflict**, including cases where Instagram sanctioned critics of vaccination or lockdown measures. The analysis also surveys legal challenges and human rights concerns, referencing **U.S. and international free speech laws**, and explores evidence of **hidden algorithms** that may systematically censor certain content for political or other motives. By compiling academic and journalistic sources, this report aims to provide a comprehensive overview of Instagram’s content moderation controversies and their implications for free expression.

Allegations of Ideological Censorship on Instagram

Accusations of **ideological bias** in Instagram’s moderation have emerged from various communities. Users with **conservative, religious, or nationalist (“patriotic”) beliefs** have frequently claimed that their content is unfairly targeted under the platform’s rules. While Instagram’s official policy forbids hate speech, misinformation, and harmful content, critics say these rules are enforced selectively to silence particular viewpoints.

- **Conservative Voices:** Prominent right-leaning politicians and commentators have long argued that social media platforms “unlawfully silence conservative viewpoints” ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)). For example, in mid-2021 former President Donald Trump filed lawsuits against Facebook/Instagram and other tech companies, alleging they violated First Amendment rights by banning him and others – though legal experts noted that private companies are not bound by the First Amendment ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)). A large majority of Republicans (about 90%) believe social media sites intentionally censor political viewpoints ([Most Americans Think Social Media Sites Censor Political Viewpoints](#)). Specific incidents fuel this belief: in August 2024, Instagram suddenly **deleted**

several accounts of prominent French right-wing activists, some with tens of thousands of followers, even though they had committed no legal offense ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#)). Those banned included leaders of nationalist or anti-government groups, such as feminist activist Alice Cordier and the whistleblower support association ASLA ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#)). Instagram cited violations (for instance, “*transphobia*” in posts critical of transgender ideology) as justification ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#)). The purge was denounced by affected users and politicians as a “*vraie purge politique*” (true political purge) and an attack on **freedom of expression**, with even French parliament members speaking out ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#)) ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#)). These cases exemplify the perception that Instagram’s moderation may disproportionately impact conservative or nationalist viewpoints.

- **Religious Beliefs:** Religious conservatives have also reported censorship. In one notable case, Instagram **removed a post by a Baptist ministry** that argued only men should be pastors, flagging it as hate speech (for “harassing” content) ([Instagram removes Baptist ministry's post on why only men should be pastors, then apologizes | U.S.](#)) ([Instagram removes Baptist ministry's post on why only men should be pastors, then apologizes | U.S.](#)). The post had quoted Biblical scripture (1 Timothy 2:12–14) to support a traditional view on gender roles in ministry. After the ministry appealed publicly, Instagram admitted it “*mistakenly removed this piece of content*” and restored the post, apologizing for the error ([Instagram removes Baptist ministry's post on why only men should be pastors, then apologizes | U.S.](#)). The incident (which the ministry described as “censoring biblical truth”) suggests that automated filters or over-zealous moderators might erroneously ban content that expresses mainstream religious positions on social issues. Similarly, in 2020 Instagram announced it would **ban content promoting “conversion therapy”** (a practice opposed by LGBTQ advocates) as a form of hate speech ([Facebook, Instagram bans content promoting 'conversion therapy'](#)). While intended to protect users, such policies have led some religious groups to complain that expressing traditional beliefs on sexuality can result in unfair censorship.
- **Patriotic and Political Content:** Self-described “patriots” and populist groups have raised alarms about shadow bans or follower suppression on Instagram. In the U.S., some conservative influencers allege that Instagram secretly limits the reach of pro-American or nationalist content, especially if it challenges prevailing political narratives. For instance, during the 2020 election season, users claimed that posts supporting then-President Trump or questioning election results were being **demoted** or hidden. (Facebook, Instagram’s parent company, did implement aggressive measures to curb election misinformation, which largely targeted narratives common in right-wing circles ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)).) Internationally, the pattern repeats: the August 2024 **mass bans of French right-wing accounts** were justified by Instagram under its policies, but critics noted none of the banned material was illegal in France ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#)). Affected users – many of whom identify as patriotic defenders of French traditions – argued that Instagram was silencing voices “*that defend France and the truth*” ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#)). Even humor was not spared; a French parody meme page “Une Bonne Droite” (literally “a good right [hook]”, a pun on right-wing) was removed despite being satirical ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#)). These examples underscore why many **ideologically right-wing or patriotic users feel systematically discriminated against** on the platform.

It should be noted that the issue of bias is hotly debated. **Studies and platform statements** sometimes contradict these claims. Facebook/Instagram’s executives have denied intentional bias (Instagram’s CEO Adam Mosseri stated in 2020 that “shadowbanning is *not a thing*” in terms of targeting specific accounts ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#))), and some external research finds no evidence of systemic anti-conservative bias ([Claim of anti-conservative bias by](#)

[social media firms is baseless ...](#)). Indeed, data from Facebook often shows that conservative pundits generate the highest engagement. Nevertheless, **perception matters** – and a Pew Research survey found the vast majority of Republican users *believe* censorship is happening ([Most Americans Think Social Media Sites Censor Political Viewpoints](#)). The following sections document concrete instances of content takedowns and account sanctions that have contributed to this perception.

Content Moderation Practices: From Bans to Shadowbans

Instagram employs a range of moderation tools that can impact users' visibility and ability to share content. These include **direct punitive actions** (like deleting posts or banning accounts) as well as **algorithmic techniques** (reducing a post's reach, or "shadowbanning"). Understanding these mechanisms is crucial to evaluating censorship claims:

- **Content Removal and Bans:** The most obvious form of moderation is deleting individual posts or permanently banning accounts that violate Instagram's Community Guidelines. For example, Instagram will remove posts containing nudity, graphic violence, explicit hate speech, or dangerous misinformation. In high-profile instances, Instagram has **banned influential accounts** for spreading false information – e.g., in February 2021, Instagram **blocked the account of Robert F. Kennedy Jr.** after repeated violations of its COVID-19 misinformation policies ([Instagram Bars RFK Jr. For Spreading Vaccine Misinformation : Coronavirus Updates : NPR](#)). A Facebook spokesperson explained that **"we removed this account for repeatedly sharing debunked claims about the coronavirus or vaccines."** ([Instagram Bars RFK Jr. For Spreading Vaccine Misinformation : Coronavirus Updates : NPR](#)). RFK Jr., a prominent vaccine skeptic, thus lost his platform for what he considered "alternative" health views, illustrating how content deemed harmful by mainstream standards is outright purged. Similarly, Instagram (as part of Meta) suspended then-President **Donald Trump's account** in January 2021 for posts allegedly inciting violence (the Capitol riot) ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)) – one of the most consequential bans, sparking intense debate on private-platform authority.
- **Demonetization and Feature Restrictions:** Beyond removal, Instagram can sanction users by turning off certain features. Accounts that repeatedly post content breaching guidelines may lose access to **monetization** (e.g. being barred from running ads, affiliate links, or other revenue features). They may also be blocked from appearing in search suggestions or Explore pages, severely cutting their audience. These measures are often less visible publicly, but creators notice the impact. For instance, numerous influencers have reported sudden drops in reach after being flagged for misinformation or policy violations, effectively losing advertising income. (Meta does not typically publicize whom it demonetizes, but its policies state that accounts violating rules can be **restricted from advertising or other revenue tools** (.))
- **"Shadowbanning" (Algorithmic Suppression):** Shadowbanning refers to **stealthily reducing a user's visibility** – their posts or comments are shown to far fewer people, or hidden entirely, without any notice to the user. Instagram long denied using true "shadowbans," but evidence indicates the platform does employ similar tactics. In one investigation, *The Markup* created test accounts and found that when certain comments were flagged (for spam or harassment), Instagram would make them invisible to all other users while still showing them to the poster – a classic shadowban symptom ([How We Investigated Shadowbanning on Instagram – The Markup](#)) ([How We Investigated Shadowbanning on Instagram – The Markup](#)). In these cases, the user receives *no notification* that their comment is suppressed. Instagram has admitted that some content **won't be recommended or will be down-ranked** if it's deemed borderline or if the account has a history of breaking rules (sometimes called "reduced distribution"). Users often infer they've been shadowbanned when their engagement drops precipitously for no obvious reason. The **Washington Post** describes shadowbanning as "a form of online censorship where you're still allowed to speak, but hardly anyone gets to hear you," implemented via "invisible algorithms" without transparency

([Shadowbanning is real: How social media decides who to silence - The Washington Post](#))

([Shadowbanning is real: How social media decides who to silence - The Washington Post](#)). This opaque practice has hit various groups: conservative activists claim their posts about topics like election fraud were quietly hidden, just as some progressive activists found posts about racial justice saw sudden declines in reach ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#)) ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#)). The lack of clarity – was it the content or the algorithm? – leaves many “victims filled with self-doubt bordering on paranoia” ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#)).

- **Hashtag and Follower Manipulation:** Instagram can also **suppress certain hashtags** or search terms, and periodic purges of fake accounts can cause users to lose followers. For example, in January 2025, users noticed that searching politically charged hashtags like “#Democrat”, “#Republicans” or “#Jan6th” returned a message that “we’ve hidden these results” ([Instagram users encounter blocked political hashtags in search results](#)) ([Instagram users encounter blocked political hashtags in search results](#)). This prompted accusations of political bias on Reddit and elsewhere, though Instagram’s spokesperson claimed it was a technical glitch affecting many terms (and promised a fix) ([Instagram users encounter blocked political hashtags in search results](#)). On the follower side, some outspoken accounts allege Instagram **forcibly unfollows** their fans to limit their reach. While Instagram hasn’t confirmed this, sudden large drops in follower counts are often explained as Instagram **culling bots or inactive accounts** from the system ([Has anyone else just lost a tone of followers in the space of ... - Reddit](#)). Those routine purges, however, can be misinterpreted by users as targeted action. The net effect of these subtler tools – hashtag blocking, algorithmic unfollowing, down-ranking – is a form of **soft censorship**: content isn’t outright banned, but it becomes much harder to find or spread.

([Instagram removes Baptist ministry's post on why only men should be pastors, then apologizes | U.S.](#))

Figure: Instagram’s logo. Critics argue the platform’s content policies are sometimes enforced in ways that disproportionately silence certain viewpoints. ([Instagram removes Baptist ministry's post on why only men should be pastors, then apologizes | U.S.](#)) ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#))

In sum, Instagram’s moderation playbook ranges from **hard removals** (deleting posts/accounts) to **soft suppression** (hiding content in algorithmic shadows). For users who experience these actions – especially without a clear explanation – it indeed feels like censorship. The following case studies illustrate how these measures were applied during recent high-profile events and controversies, often igniting debates about free speech and platform power.

Case Study 1: COVID-19 Pandemic – Health Misinformation vs. Free Speech

During the global COVID-19 pandemic (2020–2021), Instagram (like other platforms) dramatically tightened its policies to combat **misinformation** about the virus, vaccines, and public health measures. While this was driven by public health imperatives, it led to numerous clashes with individuals who posted **contrarian or skeptical views** about vaccines, lockdowns, or treatments. These users often saw Instagram’s interventions as censorship of legitimate dissent.

Enforcement Against Anti-Vaccine Content: Instagram explicitly banned many false or misleading claims about COVID-19 and vaccines, on the grounds that such misinformation could harm public health. In practice, this meant thousands of posts were removed and accounts suspended – including high-profile figures in the anti-vaccine movement. The **removal of Robert F. Kennedy Jr.’s account** in February 2021

was a landmark example. RFK Jr., a long-time vaccine critic, had amassed more than 800,000 followers by posting content sowing doubt about vaccine safety and promoting conspiracy theories (e.g. that Bill Gates was using vaccines for profit or control) ([Instagram Bars RFK Jr. For Spreading Vaccine Misinformation : Coronavirus Updates : NPR](#)) ([Instagram Bars RFK Jr. For Spreading Vaccine Misinformation : Coronavirus Updates : NPR](#)). Instagram finally banned him for repeated violations, stating that he **repeatedly shared debunked claims** that undermined trust in COVID-19 vaccines ([Instagram Bars RFK Jr. For Spreading Vaccine Misinformation : Coronavirus Updates : NPR](#)). Likewise, Instagram **shut down popular anti-vaccine hashtags** (such as #**vaccineskill** or #**vaccinescauseautism**) to stem the spread of such posts ([How anti-vaxxers get around Instagram's new hashtag controls](#)). While public health experts applauded these moves, opponents argued that tech companies were effectively **silencing a minority viewpoint** – noting that some content (e.g. anecdotes of vaccine side effects) was labeled “misinformation” despite containing true personal stories, because it contradicted official narratives.

Censorship or Safety? Instagram’s defenders contend that removing demonstrably false information (like claims that COVID vaccines implant microchips) is not censorship but responsible moderation. However, a gray area emerged around **“critics of vaccination or sanitary restrictions”** – people who, for example, questioned the efficacy of lockdowns, mask mandates, or vaccine mandates without necessarily spreading falsehoods. Many such users found their posts labeled or removed, too. Some reported that merely expressing skepticism about the speed of vaccine approvals or advocating alternative therapies led to takedowns. This created a chilling effect in which **even good-faith debate was stifled** for fear of moderation. The issue came to a head when it was revealed that **U.S. government officials had pressured social media platforms to crack down harder on COVID-19 posts**. In August 2024, Meta CEO Mark Zuckerberg disclosed in a letter that “*senior officials from the Biden Administration... repeatedly pressured our teams for months to censor certain COVID-19 content, including humor and satire*” ([Zuckerberg says Biden administration pressured Meta to censor COVID-19 content | Reuters](#)) ([Zuckerberg says Biden administration pressured Meta to censor COVID-19 content | Reuters](#)). This included White House complaints in 2021 that Facebook/Instagram were “*killing people*” by not removing enough vaccine misinformation ([Zuckerberg says Biden administration pressured Meta to censor COVID-19 content | Reuters](#)). Zuckerberg’s admission, made to Congress, implied that some content moderation decisions during the pandemic were driven not just by private policy but by government influence – a dynamic critics call **“censorship laundering”** (when governments indirectly censor via private companies to evade First Amendment scrutiny). Indeed, in 2023 a federal lawsuit (*Missouri v. Biden*) alleged that U.S. officials colluded with tech firms to suppress anti-lockdown or anti-vaccine speech, raising novel First Amendment questions. While that case is ongoing, the revelation underscores how **anti-establishment views on COVID-19 were aggressively policed on Instagram**, arguably at the expense of open discourse ([Zuckerberg says Biden administration pressured Meta to censor COVID-19 content | Reuters](#)) ([Zuckerberg says Biden administration pressured Meta to censor COVID-19 content | Reuters](#)).

Resulting Backlash: The pandemic highlighted the tension between **public safety and free expression**. Instagram likely prevented some dangerous misinformation from going viral, but its tactics also galvanized a movement decrying “Big Tech censorship.” Influencers who were deplatformed on Instagram migrated to alternative platforms (like Telegram or Parler) with cries of persecution. Even some scientists and journalists grew concerned – not because they endorsed misinformation, but because **legitimate questions or minority opinions were easily swept up in Instagram’s dragnet**. In hindsight, enforcement errors were acknowledged: Meta’s oversight board reviewed several COVID-related removal decisions, at times overturning them as unjustified. For example, the Oversight Board restored a post that had been removed for quoting a head of state’s controversial comments on COVID, noting that **discussing government policies – even if contentious – was protected political expression under international human rights standards** (). This indicates that **Instagram’s moderation, in the heat of the pandemic, sometimes overreached**. The experience has since led Meta to incorporate more transparency (like feeding some decisions through fact-checkers and offering **labels instead of removals** for borderline content). Nonetheless, the COVID-19 era stands as a case where Instagram’s fight against misinformation directly collided with individuals’ rights to voice alternative perspectives, fueling ongoing distrust among those communities.

Case Study 2: U.S. Elections and Political Speech – The Trump Era

Social media played an outsized role in the political turbulence of the late 2010s and early 2020s. Instagram, as part of Meta’s ecosystem, was involved in efforts to police election-related falsehoods and extremist content, which supporters viewed as safeguarding democracy but critics saw as **politically motivated censorship**. The saga around Donald Trump’s content and allied movements (like Stop the Steal and QAnon) illustrates this conflict.

Election Misinformation and Moderation: In the lead-up to the **2020 U.S. Presidential Election**, Instagram implemented stricter rules against false claims about voting procedures, election results, and delegitimization of outcomes. Content deemed to spread baseless allegations of widespread voter fraud – a narrative heavily pushed by Trump and supporters – was labeled or removed. For instance, hashtags associated with election conspiracy theories (e.g. **#StopTheSteal**) were blocked on Instagram and Facebook for violating rules on civic integrity. Such interventions intensified after the election: when Trump and allies continued to assert that the election was “stolen,” Instagram removed some of those posts and prevented certain phrases from trending. On January 6, 2021, as riots erupted at the U.S. Capitol, Meta took the unprecedented step of **suspending President Trump’s Facebook and Instagram accounts** indefinitely, citing the risk of further violence ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)). This ban silenced the sitting U.S. president on two of the world’s largest platforms and was widely hailed by Trump’s opponents – and equally condemned by his supporters as blatant **political censorship**.

Claims of Anti-Conservative Bias: The Trump incident became a cause célèbre for those arguing that Silicon Valley elites were biased against conservatives. Trump himself repeatedly claimed that “*right-wing voices are being censored*” on platforms like Facebook and Instagram ([Instagram users encounter blocked political hashtags in search results](#)) (despite data showing conservative content often dominates engagement ([Instagram users encounter blocked political hashtags in search results](#))). In July 2021, Trump and a group of others **filed a class-action lawsuit** against Facebook (Instagram’s parent), Twitter, and YouTube, accusing them of censoring conservatives and demanding restoration of their accounts ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)) ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)). The suit even sought the **invalidation of Section 230** of the Communications Decency Act ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)) – a U.S. law that grants platforms immunity for moderating content in good faith. Legal scholars noted the irony: Section 230 actually *protects* platforms’ ability to remove content (meaning striking it down could make companies *less* likely to host controversial speech, not more). A judge eventually dismissed Trump’s case, affirming that private companies are not bound by First Amendment free speech requirements and that no evidence turned them into “state actors” in this context ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)). However, the political pressure from the right persisted, leading states like Florida and Texas to pass laws trying to **prohibit social media censorship of political candidates or viewpoints**. (Florida’s law was blocked in court as unconstitutional, and Texas’s law is under judicial review ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#))).

Trump’s Reinstatement and Legal Victory: In a surprising development, by 2023–2024 Meta began reconsidering some bans. Under internal policies (and guidance from the independent Oversight Board), they announced Trump’s suspension would be lifted after two years (i.e., in 2023) with “guardrails” on his future posts. Then in **January 2025, Meta agreed to settle Trump’s lawsuit** before it went any further – paying **\$25 million** to Trump to resolve his claims that Facebook/Instagram had engaged in “impermissible censorship” by banning him ([Meta agrees to pay Trump \\$25 million to settle lawsuit over Facebook and Instagram suspensions : NPR](#)). This settlement, reportedly with a large sum directed to Trump’s Presidential Library, was seen as a “*significant victory for Trump*” and raised eyebrows across the political spectrum ([Meta agrees to pay Trump \\$25 million to settle lawsuit over Facebook and Instagram suspensions : NPR](#)). Some critics (like Senator Elizabeth Warren) blasted the payout as “*a bribe... signal[ing] that corruption is the name of the game*,” suggesting Meta might expect political favors in return ([Meta agrees to pay Trump \\$25 million to settle lawsuit over Facebook and Instagram suspensions : NPR](#)). Regardless of motivations, the settlement effectively ended the dispute with Trump without a court precedent, but it acknowledged the **political reality that Meta needed to mend fences with the right**. By late 2024, Mark Zuckerberg even

ordered a softening of some moderation rules – for example, **dropping strict fact-checking in favor of “community notes” and allowing political content to be recommended again** – in what many saw as an attempt to counter claims of anti-conservative bias as the 2024 election approached ([Instagram users encounter blocked political hashtags in search results](#)) ([Instagram users encounter blocked political hashtags in search results](#)). The result has been a subtle shift: Instagram is comparatively less heavy-handed with political speech now than it was in 2020, illustrating how public and legal pushback can influence platform policies.

In summary, Instagram’s handling of Trump-era political content reveals a complex interplay of **principle, pressure, and profit**. Initially, the platform took a hard line on what it viewed as dangerous falsehoods (culminating in Trump’s ban), aligning with many users’ view that moderation was necessary. Yet the backlash from conservatives and the threat of regulatory/legal action forced a re-evaluation. The episode underscores why debates over social media moderation are often **polarized along partisan lines** – with one side demanding more enforcement against “misinformation” and the other side calling it “censorship” of their worldview. Instagram found itself squarely in the middle of that national argument.

Case Study 3: Geopolitical Conflicts and Censorship – Ukraine and Beyond

Instagram’s moderation challenges are not limited to U.S. politics or health debates; they also extend to international conflicts, where the platform’s policies sometimes appear to favor one side’s narrative, raising concerns of censorship on a global scale. Two significant contexts are the **Russia-Ukraine war (2022–present)** and the **Israel-Palestine conflict (2023)**, both of which saw unusual content moderation decisions by Instagram that drew criticism.

Ukraine War – Partiality in Policy: In March 2022, shortly after Russia’s invasion of Ukraine, internal Meta emails (later confirmed to Reuters) revealed a startling temporary policy change: Facebook and Instagram decided to **allow users in certain countries to post calls for violence against Russian soldiers and leaders** in the context of the war ([Facebook allows war posts urging violence against Russian invaders | Reuters](#)) ([Facebook allows war posts urging violence against Russian invaders | Reuters](#)). Normally, statements like “death to [individuals]” violate hate speech or incitement rules. But Meta made an exception, **permitting statements such as “death to Russian invaders”** or even calls for the death of President Putin, as long as they weren’t credible threats ([Facebook allows war posts urging violence against Russian invaders | Reuters](#)). This was framed as letting Ukrainians and their neighbors express their anger at an aggressor – essentially taking an ideological stance that supported Ukraine’s defense. The fallout was immediate: Russia’s government blasted Meta for “extremist activities” ([Facebook allows war posts urging violence against Russian invaders | Reuters](#)) and within days **banned Instagram entirely in Russia**, cutting off tens of millions of users ([Russia bans Facebook and Instagram under 'extremism' law](#)) ([Russia is restricting social media as war in Ukraine continues - NPR](#)). From one perspective, Instagram’s allowance was a show of solidarity with a nation under attack; from another, it showed a **double standard in content moderation** – hate speech was tolerated when directed at a politically disfavored group (Russian military/leadership), something that would be unthinkable in other contexts. Conversely, any content from the Russian side that justified the war or was seen as propaganda was aggressively down-ranked or removed by Meta, especially after Western governments and the EU called for clamping down on Russian state media. Instagram labeled state-affiliated Russian media accounts and, in line with EU sanctions, **blocked outlets like RT and Sputnik in EU countries** (a form of enforced censorship of those sources).

Hidden Censorship in Conflict-Related Content: Beyond overt policy changes, Instagram’s algorithms quietly struggled with war-related content. A 2024 investigation by *The Markup* found that Instagram **“heavily demoted” even non-graphic images of war**, deleted some post captions without warning, and hid

numerous comments – often without notifying users ([Demoted, Deleted, and Denied: There’s More Than Just Shadowbanning on Instagram – The Markup](#)). It also **suppressed hashtags** in inconsistent ways. Notably, users who tried to post about the war (for instance, sharing images or reports from Ukraine) sometimes found their posts inexplicably getting far fewer views, implying the algorithm flagged them. In some cases, comments about the war were removed and tagged as “spam” (thus also *denying the user an option to appeal*, since Instagram doesn’t allow appeals on spam-classified removals) ([Demoted, Deleted, and Denied: There’s More Than Just Shadowbanning on Instagram – The Markup](#)). This affected commentary on both sides, but activists complained that important documentation – such as Ukrainian civilians’ posts about atrocities – might have been buried by automated filters not tuned for a war scenario ([Designed for peacetime, not war: How Ukraine is forcing companies ...](#)). Instagram’s moderation systems were arguably “**designed for peacetime, not war,**” leading to overzealous filtering of conflict content ([Designed for peacetime, not war: How Ukraine is forcing companies ...](#)). The company had to adjust some algorithms after Ukrainian government criticism that platforms were removing evidence of Russian war crimes under mistaken content rules (e.g., graphic violence policies). This scenario highlights that **censorship can occur not only from deliberate bias but from poorly calibrated algorithms** that err on the side of removal during crises.

Israel-Palestine Conflict – One-Sided Silencing: Another prominent example occurred during the Israel–Hammas war of October 2023. Multiple reports indicate Instagram and Facebook engaged in **systematic takedowns of content supportive of Palestinians** or critical of Israeli military actions. Human Rights Watch investigated and documented over **1,000 cases of content removals or suppressions** between October and November 2023, “*of which 1,049 involved peaceful content in support of Palestine ... and only 1 involved content in support of Israel.*” ([Meta’s Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook | HRW](#)). This included posts about human rights abuses, protests, or sympathy for Gaza civilians, which were removed or hidden at vastly disproportionate rates ([Meta’s Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook | HRW](#)) ([Meta’s Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook | HRW](#)). HRW called this a “*wave of heightened censorship*” that “**silenced voices in support of Palestine**” globally ([Meta’s Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook | HRW](#)). One cause was likely aggressive enforcement of rules against praising terrorist organizations – Hamas is banned on Meta – but the net cast by AI moderation was far too wide, catching benign speech by Palestinians themselves about their suffering. Instagram also “**erratically suppressed**” **hashtags** like #FreePalestine or #CeasefireForGaza according to user reports, and often labeled posts as sensitive or demoted them, making them less visible ([Demoted, Deleted, and Denied: There’s More Than Just Shadowbanning on Instagram – The Markup](#)). The **public outcry** from journalists, activists, and even some Meta employees led the Oversight Board to announce a case on this issue. Meta responded by blaming some problems on bugs (for instance, a glitch that temporarily caused Stories about the war to appear lower in feeds) and promising to fix errors. Nonetheless, the perception remains that Instagram **took sides – intentionally or not – by muzzling one side’s narrative** in an international conflict. This not only damages the platform’s neutrality but, as HRW noted, *breaches the responsibility to respect freedom of expression* under international human rights standards ([Meta’s Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook | HRW](#)).

In these geopolitical cases, Instagram’s moderation was tested against fast-moving, emotionally charged content. The platform’s actions, whether **policy-driven (Ukraine)** or **error-driven (Palestine)**, ended up raising allegations of censorship and bias. They show that Instagram’s “**official ideology**” (which aligns with Western liberal perspectives and compliance with Western government norms) can come into conflict with alternative viewpoints – be it a Russian patriot defending his country’s actions, or a Palestinian voice highlighting their plight. When those views are suppressed, it fuels arguments that Instagram violates the spirit of free expression beyond U.S. borders, even if legally the First Amendment doesn’t apply to a private company.

Legal and Human Rights Considerations

The tension between Instagram’s moderation practices and users’ speech rights has prompted reference to legal frameworks, both American and international, that protect freedom of expression. **At the core, however, is a paradox:** Instagram is a private corporation, and thus **not directly bound by the First Amendment** of the U.S. Constitution (which restrains only government censorship) ([Free Speech on Social Media: The Complete Guide - Freedom Forum](#)). In fact, as U.S. courts have affirmed, social media companies themselves have a First Amendment right to decide what speech to allow on their platforms – essentially their own freedom of expression as private actors ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)). This means that users cannot successfully sue Instagram for violating their constitutional free speech rights, as Trump’s dismissed lawsuit showed.

Nonetheless, **legal and regulatory battles** are being fought to define the boundaries of platform power:

- **United States Law:** While the First Amendment doesn’t apply to Instagram’s content decisions, other laws and doctrines come into play. **Section 230 of the Communications Decency Act (47 U.S.C. § 230)** is particularly significant. Section 230 provides that online platforms aren’t liable for content posted by users, and importantly, they aren’t liable for “**good faith**” **content removal** either. This law, termed the “backbone of the internet,” has allowed Instagram to remove or restrict content it deems objectionable without fear of being sued for bias ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)). Critics of Section 230 (including some conservatives) argue it gives tech companies too much power to censor, and they have lobbied to modify or repeal it ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)). So far, Section 230 remains in force, and attempts to use legal means to force Instagram to carry certain speech (such as Florida’s social media law penalizing deplatforming of political candidates) have run into constitutional problems, as courts ruled that government cannot compel a private platform to host speech either ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)). Thus, under current U.S. law, Instagram has broad leeway to moderate content as it sees fit – but this very freedom is what some state legislatures and politicians are challenging.
- **International Human Rights Law:** Free expression is also protected by international instruments that set global norms. **Article 19 of the Universal Declaration of Human Rights (1948)** proclaims the right to “*seek, receive and impart information and ideas through any media and regardless of frontiers.*” Similarly, **Article 19 of the International Covenant on Civil and Political Rights (ICCPR)** (a binding treaty ratified by the U.S.) guarantees everyone the right to freedom of expression, subject only to narrow restrictions (such as incitement to violence or national security concerns) ([What is freedom of expression? - ARTICLE 19](#)). These provisions, however, apply to **governments** – they obligate states to not censor unlawfully. Instagram, as a private sector entity, is not directly a duty-bearer under these treaties. However, there is a growing expectation that companies will **voluntarily uphold human rights standards**. The **UN Guiding Principles on Business and Human Rights** call on corporations to respect human rights in their operations, and Meta has indeed adopted a **corporate human rights policy** pledging commitment to principles like freedom of expression (). In practice, Meta’s **Oversight Board** explicitly uses international human rights law as a benchmark when reviewing difficult content moderation cases. For instance, when evaluating whether a removal was proper, the Board asks if the decision was “*necessary and proportionate*” under Article 19 ICCPR standards (). In some cases, the Board has found Instagram’s actions inconsistent with human rights—for example, removing a post about a **political prisoner (Abdullah Öcalan)** was deemed unnecessary and disproportionate, as the content was political advocacy posing no direct harm (). Such analysis mirrors what a court might do if ICCPR Article 19 bound Instagram.
- **Notable Legal Cases:** Besides Trump’s lawsuit, other legal or quasi-legal proceedings have scrutinized Instagram’s moderation. In Russia, as noted, a court branded Meta an “extremist” organization partly because of its moderation choices, effectively outlawing Instagram nationwide ([Russia bans Facebook and Instagram under 'extremism' law](#)). In the EU, the new **Digital Services Act (DSA)** requires major platforms to be transparent about their algorithms and content removal decisions, and to offer users appeals – a legal push for due process in moderation. While not directly about free speech, laws like the DSA and various proposed regulations aim to **check the arbitrary power** of platforms like Instagram by mandating audits of their systems (which could expose any

political biases or censorship directives in their algorithms). We also see lawsuits from individuals: e.g., some conservative influencers have sued for defamation or discrimination when their content was labeled misinformation (arguing the labels damaged their reputation or business). These cases face uphill battles but indicate a trend of using creative legal arguments to challenge Instagram's content actions.

In essence, **American law currently sides with Instagram's right to curate content**, whereas **international human rights principles encourage Instagram to align its policies with free expression norms**. There is an inherent conflict: Instagram can legally remove speech that is lawful (since it's a private forum), but doing so excessively or unequally can violate the spirit of human rights. As Human Rights Watch observed in the context of Palestinian content takedowns, Meta's inconsistent enforcement led to "*erroneous removal of content*" and **disproportionate silencing** of one viewpoint ([Meta's Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook | HRW](#)), undermining users' rights. To address such concerns, experts recommend that Instagram **adopt clearer rules, ensure proportionality in enforcement, and improve transparency and appeals** (). Some of these steps are gradually being implemented under public and regulatory pressure.

Hidden Algorithms, Shadow Bans, and Transparency Concerns

Perhaps the most insidious form of censorship alleged on Instagram is not the **overt removal** of content, but the **covert algorithmic throttling** of it. Instagram's feed algorithm and content-ranking systems can effectively make a post invisible without the user knowing – a phenomenon popularly known as **shadow banning**. The lack of transparency around these algorithms has led to widespread concern about "**invisible**" **censorship** for political or even corrupt motives (such as favoring content that aligns with Instagram's business interests or external pressures).

Understanding the Algorithmic Black Box: Instagram uses algorithms to curate what users see – both in their main feed and in the Explore tab or hashtag searches. These algorithms are tuned to maximize user engagement and safety, which means they don't simply show everything from everyone a user follows in chronological order. By design, they **down-rank content** that is likely to be reported or that the system suspects might be problematic (even if it hasn't been removed outright). For example, if a post is flagged by Instagram's AI as potential misinformation, its distribution might be reduced by 80%, even if the post remains visible on the author's profile. This creates scenarios where a user's content is technically allowed on Instagram but **hardly reaches anyone**, which is effectively a form of censorship. It's "shadowy" because the user is often not notified – their post isn't deleted, so they receive no warning, yet the impact (drastically reduced audience) is real ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#)) ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#)).

Investigations like *The Markup's "Automated Censorship" series* have shed light on these practices. The Markup found that Instagram uses different labels (spam, bullying, etc.) in ways that affect a user's ability to appeal and their content's visibility. In tests, comments that Instagram tagged as "**spam**" led to **instant shadowbans** – the commenter could see their remark, but no one else could – whereas comments tagged as **"harassment"** allowed the user an option to appeal the removal ([How We Investigated Shadowbanning on Instagram – The Markup](#)) ([How We Investigated Shadowbanning on Instagram – The Markup](#)). This suggests an algorithmic strategy: by classifying certain political comments under a no-appeal category like spam, the platform can remove them quietly and preclude pushback. Indeed, during the 2023 Israel-Palestine content removals, many pro-Palestinian comments were **mysteriously removed as "spam"**, preventing users from contesting the decision ([How We Investigated Shadowbanning on Instagram – The Markup](#)) ([How We Investigated Shadowbanning on Instagram – The Markup](#)). This **opaque use of algorithms** can be seen as a deliberate (if indirect) method to censor contentious viewpoints while minimizing backlash.

[\(Demoted, Deleted, and Denied: There's More Than Just Shadowbanning on Instagram – The Markup\)](#)

Figure: Illustration representing the concept of “shadow banning” on social media. Platforms like Instagram can make a user virtually invisible by demoting or hiding their posts algorithmically.

[\(Shadowbanning is real: How social media decides who to silence - The Washington Post\)](#) [\(Shadowbanning is real: How social media decides who to silence - The Washington Post\)](#)

Political or Corrupt Motives: There is ongoing debate about whether Instagram’s hidden censorship is politically motivated or driven by other factors (like advertiser pressure or user reports). Some critics suggest that Instagram’s parent company, Meta, may tune algorithms to appease powerful governments or interest groups. For instance, as noted earlier, Mark Zuckerberg’s conciliatory moves ahead of Trump’s 2024 campaign (restoring features that boost political content) came after explicit threats and pressure from Republicans ([Instagram users encounter blocked political hashtags in search results](#)) ([Instagram users encounter blocked political hashtags in search results](#)). This raises the question: were algorithms previously tweaked to *reduce* the visibility of right-leaning political content, and then adjusted back when political winds changed? While Meta denies partisan manipulation, the timing of policy shifts can appear suspect. On the flip side, human rights advocates accuse Instagram of succumbing to **government requests that border on censorship** – e.g., temporarily hiding hashtags critical of certain regimes or events at the behest of authorities. (One example outside our main scope: reports that Instagram limited reach of #EndSARS during Nigerian protests, after government complaints, though the company cited a technical error.) **Corruption** could also entail preferential treatment: whistleblower documents (the Facebook Files) revealed an internal program called “XCheck” that exempted millions of VIP users from standard moderation. While XCheck was meant to avoid false positives on important accounts, it effectively meant some elites could spread content (including misinformation) that ordinary users would be penalized for (). This unequal application of rules undermines trust – it suggests the algorithm can be **selective and unprincipled**, harsher on some users, lenient on others, for reasons unrelated to the content itself.

Calls for Transparency and Reform: The emerging consensus among experts is that **sunlight is the best remedy** for concerns about hidden censorship. Transparency reports, independent audits, and user empowerment tools are being advocated. Users want to know, as *The Washington Post* put it, “*when and why [the platforms] are suppressing our megaphones*” ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#)). In response, Instagram has made some moves: it now periodically publishes “**Widely Viewed Content**” reports and has an account status feature that tells users if their posts are being recommended or not. But these are baby steps. The Digital Services Act in Europe will force Instagram to allow vetted researchers to study its algorithms for systemic biases. Additionally, Instagram’s **Oversight Board** has urged clearer definitions – for example, telling Meta to define what constitutes “borderline” content that gets down-ranked and to notify users when their reach is reduced.

Without such measures, the suspicion remains that **hidden algorithms can be weaponized for censorship**. Whether it’s an engineer’s conscious tweak or simply the emergent behavior of AI moderation systems, the effect is that certain views (often those challenging a status quo or “official” position) struggle to find an audience on Instagram. In authoritarian countries, this dynamic is even more concerning: if Meta can dial down content to suit a government (or must do so to avoid being banned entirely), then user speech rights are at the mercy of backroom algorithmic adjustments. Going forward, the push for algorithmic accountability is critical. As one survey found, *nearly 1 in 10 U.S. social media users believe they have been shadowbanned* – real or not, that perception erodes trust ([Shedding Light on Shadowbanning](#)). Only through **greater transparency, consistent policy application, and avenues for redress** can Instagram dispel the notion that it operates a secret censorship machine.

Conclusion

The examination above reveals that Instagram’s approach to content moderation has repeatedly brought it into conflict with principles of free expression and has led to allegations of **ensorship and discrimination** against users espousing “alternative” views. From conservative commentators and religious groups to activists in conflict zones, a wide array of users have experienced removals, bans, or shadow bans that they perceive as unjust. In some cases, Instagram did overreach – as shown when it **mistakenly removed harmless content** (later restoring it) ([Instagram removes Baptist ministry's post on why only men should be pastors, then apologizes | U.S.](#)), or when investigations found **systematic bias** in enforcement (e.g., disproportionately censoring pro-Palestinian content) ([Meta’s Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook | HRW](#)). In other cases, the platform was arguably enforcing its rules against hate speech or misinformation legitimately, yet the **impacts were felt along ideological lines**, feeding a narrative of deliberate silencing. The COVID-19 pandemic and the 2020 U.S. election, especially, showcased Instagram’s willingness to side with what it viewed as truth against falsehood – even at the cost of alienating those who held the opposing opinions. This has raised profound questions: *Should a private company of Instagram’s reach have the power to decide which political or scientific ideas are acceptable?* And if it does wield that power, *what checks should exist to prevent abuse or mistakes?*

Legally, Instagram is within its rights to moderate content (protected by laws like Section 230 in the U.S.), but the **court of public opinion and international human rights standards** hold it to a higher standard. **Freedom of speech** is cherished globally as a fundamental right, and while Instagram is not a government, its platform has become a de-facto public square for discourse. When users are blocked or shadowbanned, their ability to participate in that public square is impeded, raising human rights concerns. Documents and testimonies (such as Zuckerberg’s letter to Congress) have even suggested that at times **Instagram’s censorship was encouraged by government officials** ([Zuckerberg says Biden administration pressured Meta to censor COVID-19 content | Reuters](#)) ([Zuckerberg says Biden administration pressured Meta to censor COVID-19 content | Reuters](#)), blurring the line between private content moderation and state censorship-by-proxy.

To mitigate these issues, various solutions are being discussed. These include: **clearer Community Guidelines**, applied evenly across ideologies; improved **appeal processes and transparency**, so users know why action was taken and can contest it (no more unexplained “spam” labels on political speech); **regular audits** of algorithmic outcomes to detect unintended bias; and possibly **third-party oversight** to review contentious moderation decisions (expanding on the Oversight Board model). On the user side, there is a push for **digital rights awareness**, reminding people that while free speech is a right, private platforms also have rights – hence the importance of creating alternative outlets and not relying on a single corporate forum for one’s voice.

Instagram stands at a crossroads common to big tech: it must balance **safety, truth, and corporate values** with the **messy, sometimes offensive realm of free expression**. Its history of content disputes – from conservatives’ complaints to international incidents – shows that erring too far on either side invites backlash. In the coming years, the hope is that through greater transparency and adherence to principled, **human-rights-based policies**, Instagram can reduce perceptions of viewpoint discrimination. Ensuring that users of all stripes feel their voice can be heard (within the bounds of law and respect) is essential for the platform’s legitimacy. The challenge is formidable: to create an online environment that is both **civil and open**, where **harmful disinformation is curbed** but **healthy debate – including dissenting, alternative views – is allowed to thrive**. Whether Instagram can rise to that challenge will significantly influence the global discourse on social media and free speech in the digital age.

Sources:

1. Christian Post – Instagram apologizes for removing Baptist ministry post (2018) ([Instagram removes Baptist ministry's post on why only men should be pastors, then apologizes | U.S.](#)) ([Instagram removes Baptist ministry's post on why only men should be pastors, then apologizes | U.S.](#))
2. European Conservative – Instagram bans French right-wing accounts (Aug 2024) ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#)) ([Instagram Shuts Down Accounts of Several French Right-Wing Voices — The European Conservative](#))

3. Reuters – Zuckerberg letter: White House pressured Meta to censor COVID content ([Zuckerberg says Biden administration pressured Meta to censor COVID-19 content | Reuters](#)) ([Zuckerberg says Biden administration pressured Meta to censor COVID-19 content | Reuters](#))
4. Reuters – Meta allows calls for violence against Russian invaders (Mar 2022) ([Facebook allows war posts urging violence against Russian invaders | Reuters](#)) ([Facebook allows war posts urging violence against Russian invaders | Reuters](#))
5. NPR – Meta settles with Trump over Facebook/Instagram ban, pays \$25M (Jan 2025) ([Meta agrees to pay Trump \\$25 million to settle lawsuit over Facebook and Instagram suspensions : NPR](#)) ([Meta agrees to pay Trump \\$25 million to settle lawsuit over Facebook and Instagram suspensions : NPR](#))
6. Reuters – Trump sues social media alleging bias, Section 230 challenge (Jul 2021) ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#)) ([Trump sues Facebook, Twitter and Google, claiming censorship | Reuters](#))
7. Axios – Instagram blocked political hashtags, users allege bias (Jan 2025) ([Instagram users encounter blocked political hashtags in search results](#)) ([Instagram users encounter blocked political hashtags in search results](#))
8. Washington Post – Shadowbanning on social media and calls for transparency ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#)) ([Shadowbanning is real: How social media decides who to silence - The Washington Post](#))
9. *The Markup* – Investigation finds Instagram demoted war content, hid comments (Feb 2024) ([Demoted, Deleted, and Denied: There's More Than Just Shadowbanning on Instagram – The Markup](#))
10. NPR – Instagram bans Robert F. Kennedy Jr. for vaccine misinformation ([Instagram Bars RFK Jr. For Spreading Vaccine Misinformation : Coronavirus Updates : NPR](#)) ([Instagram Bars RFK Jr. For Spreading Vaccine Misinformation : Coronavirus Updates : NPR](#))
11. Human Rights Watch – Meta's censorship of pro-Palestine content (Dec 2023) ([Meta's Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook | HRW](#))

YOUR DONATION EMPOWERS OUR CAUSE

HUMAN RIGHTS & ANALYTICAL HOUSE, INC.



zelle



PayPal
VISA & MasterCard